REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
12Jun2014	Final Report	11Dec2009-31Dec2013

4. TITLE AND SUBTITLE
Developing cDNA Libraries of Receptors Involved in the Recruitment of the Biofouling Tubeworm Hydroides elegans

5a. CONTRACT NUMBER

5b. GRANT NUMBER
N00014-08-1-0413

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Hadfield, M.G.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Hawaii at Manoa
41 Ahui Street
Honolulu, HI 96813

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Department of Defense
Office of Naval Research

10. SPONSOR/MONITOR'S ACRONYM(S)
DOD, ONR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

13. SUPPLEMENTARY NOTES

14. ABSTRACT
See Attached

15. SUBJECT TERMS
Transcriptome, Hydroides elegans, Next Generation Sequencing, Illumina HiSeq, PacBio SMRT, Biofilm, Metamorphosis

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Michael G. Hadfield
U	U	U	UU	14	19b. TELEPHONE NUMBER *(Include area code)* 808-539-7319

INSTRUCTIONS FOR COMPLETING SF 298

1. REPORT DATE. Full publication date, including day, month, if available. Must cite at least the year and be Year 2000 compliant, e.g. 30-06-1998; xx-06-1998; xx-xx-1998.

2. REPORT TYPE. State the type of report, such as final, technical, interim, memorandum, master's thesis, progress, quarterly, research, special, group study, etc.

3. DATE COVERED. Indicate the time during which the work was performed and the report was written, e.g., Jun 1997 - Jun 1998; 1-10 Jun 1996; May - Nov 1998; Nov 1998.

4. TITLE. Enter title and subtitle with volume number and part number, if applicable. On classified documents, enter the title classification in parentheses.

5a. CONTRACT NUMBER. Enter all contract numbers as they appear in the report, e.g. F33315-86-C-5169.

5b. GRANT NUMBER. Enter all grant numbers as they appear in the report. e.g. AFOSR-82-1234.

5c. PROGRAM ELEMENT NUMBER. Enter all program element numbers as they appear in the report, e.g. 61101A.

5e. TASK NUMBER. Enter all task numbers as they appear in the report, e.g. 05; RF0330201; T4112.

5f. WORK UNIT NUMBER. Enter all work unit numbers as they appear in the report, e.g. 001; AFAPL30480105.

6. AUTHOR(S). Enter name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. The form of entry is the last name, first name, middle initial, and additional qualifiers separated by commas, e.g. Smith, Richard, J, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES). Self-explanatory.

8. PERFORMING ORGANIZATION REPORT NUMBER. Enter all unique alphanumeric report numbers assigned by the performing organization, e.g. BRL-1234; AFWL-TR-85-4017-Vol-21-PT-2.

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES). Enter the name and address of the organization(s) financially responsible for and monitoring the work.

10. SPONSOR/MONITOR'S ACRONYM(S). Enter, if available, e.g. BRL, ARDEC, NADC.

11. SPONSOR/MONITOR'S REPORT NUMBER(S). Enter report number as assigned by the sponsoring/ monitoring agency, if available, e.g. BRL-TR-829; -215.

12. DISTRIBUTION/AVAILABILITY STATEMENT. Use agency-mandated availability statements to indicate the public availability or distribution limitations of the report. If additional limitations/ restrictions or special markings are indicated, follow agency authorization procedures, e.g. RD/FRD, PROPIN, ITAR, etc. Include copyright information.

13. SUPPLEMENTARY NOTES. Enter information not included elsewhere such as: prepared in cooperation with; translation of; report supersedes; old edition number, etc.

14. ABSTRACT. A brief (approximately 200 words) factual summary of the most significant information.

15. SUBJECT TERMS. Key words or phrases identifying major concepts in the report.

16. SECURITY CLASSIFICATION. Enter security classification in accordance with security classification regulations, e.g. U, C, S, etc. If this form contains classified information, stamp classification level on the top and bottom of this page.

17. LIMITATION OF ABSTRACT. This block must be completed to assign a distribution limitation to the abstract. Enter UU (Unclassified Unlimited) or SAR (Same as Report). An entry in this block is necessary if the abstract is to be limited.

ABSTRACT

The serpulid polychaete, *Hydroides elegans* is a problem fouler in warm water ports throughout the world and is an excellent candidate for investigating molecular reception during larval settlement and metamorphosis. *H. elegans* is induced to settle and metamorphose when given the appropriate bacterial cue. This study utilized transcriptome sequencing of *H. elegans* and genome sequencing of two inductive bacterial strains (i.e., *Pseudoalteromonas luteoviolacea* and *Cellulophaga lytica*) to provide a means to examine mechanisms and/or genes that may be significant in understanding the interaction between bacteria and larvae that result in biofouling. Transcriptome analysis and *in-situ* hybridization techniques determined the genetic expression and location of genes in *H. elegans* as well as important biological pathways involved in metamorphic competence. Likewise, the genome sequencing of *P. luteoviolacea* and *C. lytica* determined significant coding sequences for structures that play a role in the settlement and metamorphosis of *H. elegans*.

FINAL TECHNICAL REPORT

Review Period: June 1, 2013-December 31, 2013

TITLE: Developing cDNA Libraries of Receptors Involved in the Recruitment of the Biofouling Tubeworm *Hydroides elegans*

ONR AWARD NUMBER: N00014-08-1-0413

PRINCIPLE INVESTIGATOR: Michael G. Hadfield, Ph.D.
University of Hawaii at Manoa
Kewalo Marine Laboratory
41 Ahui Street
Honolulu, HI 96813

Name and Address of Submitting Organization: University of Hawaii at Manoa
Sakamaki Hall D-200
2530 Dole Street
Honolulu, HI 96822

I. Scientific and Technical Objectives

The tubeworm *Hydroides elegans*, is a significant problem fouler in warm water ports around the world and is an excellent candidate for investigating molecular reception during larval settlement and metamorphosis (Hadfield et al. 1994). Modern molecular techniques enable us to examine and identify larval receptor systems providing a significant contribution toward understanding the interaction between bacterial ligands and larval receptors that result in biofouling. Expanding upon the research funded by ONR Grant No. N00014-08-1-0413, we will determine the molecular basis of bacterial induction of larval metamorphosis and identify biological pathways and genes expressed at the stage of metamorphic competence of the larvae of *H. elegans*. Additionally, *H. elegans* is induced to settle and metamorphose as a response to a bacterial cue from a bacterial biofilm. Recently, this cue has been identified to be a phage-tail like bacteriocin produced by the bacterium *Pseudoalteromonas luteoviolacea* (Shikuma et al., 2014). However, in addition to *P. luteoviolacea*, *Cellulphaga lytica* has also been identified as a bacterium capable of inducing the settlement and metamorphosis of *H. elegans*. Therefore, sequencing the genome of these inductive bacterial species allows discovery of bacterial products and/or metabolites that may be involved in the induction of larval settlement and metamorphosis – the fundamental basis of marine biofouling.

II. Approach

A. Transcriptome Sequencing of *Hydroides elegans*

The initial approach to screening *H. elegans* for candidate receptor genes involved a large arrayed cDNA printed library developed in the laboratory of Dr. Andrew Cameron at the California Institute of Technology. Approximately 200,000 cDNA gene fragments that were cloned in bacteria were contained on a filter to be screened for possible receptor genes. In this process, degenerate primers were created from sequences of candidate genes that were not necessarily specific to *H. elegans* but instead found in the GenBank database of other marine invertebrates. Recently, however, more advanced and accurate methods for "mining" for genes of interest have developed through Next Generation Sequencing (NGS) technology. One such method used in NGS is the Illumina HiSeq method. Illumina HiSeq allows for the generation of multiple genetic sequences (i.e., reads) from a biological sample. As a result, the subsequent approach taken by this laboratory to determine the potential receptor sequences and expression of specific proteins involved in the settlement and metamorphosis of *H. elegans*, was to sequence the transcriptome of *H. elegans*. The transcriptome of the larval (precompetent and competent) and adult stages of *H. elegans* was determined since the transcriptome of any organism reflects the genes that are being actively expressed at any given time.

Specifically, total RNA samples were prepared using the Qiagen RNeasy Mini Kit and the RNA quality was determined with the Agilent Bioanalyzer. The total RNA sample was then submitted to the Huntsman Cancer Institute at the University of Utah and the subsequent isolation of mRNA was used for Illumina HiSeq 101 paired end sequencing. Sequencing for all larval and adult stages was run in triplicate (biological replicate) and in individual lanes (technical replicates) of the flow cell. The transcriptome sequences generated with the Illumina HiSeq

method were aligned, annotated and assembled with specific software programs. Additionally, those genes that were differentially expressed in precompetent, competent and adult stages were determined through bioinformatic analysis. Primers were designed to the transcriptome sequences to amplify potential receptor genes and then used in the following capacities: 1) to produce a 700-900 basepair fragment and 2) to use the PCR product in *in-situ* hybridization studies. The riboprobes that are generated for the *in-situ* hybridization studies were then used to assess the anatomical location(s) of the transcripts in the competent larvae of *H. elegans* using the Roche Dig/α-Dig labeling system and visualized with a fluorescent microscope. Additionally, the genes that were differentially expressed and annotated in competent larvae were then imported into the program DAVID (**D**atabase for **A**nnotation, **V**isualization and **I**ntegrated **D**iscovery) to determine the biological pathways and genes involved at the stage of metamorphic competence.

B. Genomic Sequencing of Inductive Bacterial Strains

A second approach involved the genomic sequencing of two bacterial genera that are capable of inducing the settlement and metamorphosis of *H. elegans*. The first, more widely studied marine bacterium is *Pseudoalteromonas luteoviolacea* (HI1) and the second is *Cellulophaga lytica* (HI1).

Both *P. luteoviolacea* (HI1) and *C. lytica* (HI1) were originally isolated from a seawater table at the Kewalo Marine Laboratory (Honolulu, HI). The DNA was extracted from overnight cultures using the MoBio UltraClean Microbial DNA Isolation Kit and DNA quality and quantity were determined using the Nanodrop 2000 platform (Thermo Scientific). The isolated DNA was then submitted to the National Center for Genome Resources (NCGR) for PacBio Single Molecule Real Time (SMRT) Sequencing, a third generation sequencing method.

III. Concise Accomplishments

A. Transcriptomes of *Hydroides elegans*

Approximately 100 million reads were generated by the Illumina HiSeq 101 paired end sequencing method for each developmental stage of *H. elegans*. The reads were then aligned using the software program Burrows Wheeler Aligner (BWA). Once the reads were aligned with BWA, the de novo reconstruction of the transcriptomes from the RNA-Seq data was accomplished by the Trinity software program. Each sequence was then submitted to a Blast search against various repositories of protein databases found at NCBI, Swiss-Prot, TrEMBL and NR (Non-Redundant database). Also, those sequences that were unique to each stage were determined using EBSeq, a software program for differential expression analysis and categorized into different patterns (i.e., Pattern 1-5) (Table 1).

The initial use of the transcriptome sequencing data was to design riboprobes that targeted genes related to the innate immune system since *H. elegans* responds to bacterial cues and the innate immune reaction is the first line of defense for animals against various pathogens (i.e., bacteria, viruses, etc.). Thus, the riboprobes were designed targeting the innate immune receptors;

Galactose Binding Lectin (GBL), Peptidoglycan Recognition Protein (PGRP) and NF Kappa B. GBL is a C-type lectin receptor that recognizes galactose in bacterial LPS. PGRP is a highly conserved pattern recognition receptor that binds to components of the bacterial cell wall and can activate toll receptors as part of the innate immune response. NF-kappa-B is a protein complex that controls the transcription of DNA (transcription factor) in response to stimuli such as stress and/or viral or bacterial antigens. The riboprobes were used in *in-situ* hybridization experiments to locate and visualize the anatomical location of the transcripts on the competent larvae of *H. elegans*.

Additionally, as mentioned previously, the genes that were differentially expressed in competent larvae were determined (i.e., Pattern 3) (Table 1) with the software EBSeq and analyzed using DAVID. Of the 6,422 coding sequences expressed in competent larvae, 4,679 contigs were associated with peptides of known proteins. Of the 4,679 peptides, 2031 peptides were unique to competent larvae. DAVID analysis was able to further classify the 2,031 competent larvae gene expression patterns into KEGG Pathways and Gene Ontologies (GO). The KEGG Pathway was designed to link genes in the genome to gene products in a pathway to see which pathways and associated functions are likely to be encoded in the genome of the organism. KEGG pathways are classified into sections which include 1) Metabolism 2) Genetic Information Processing 3) Environmental Information Processing 4) Cellular Processes 5) Organismal systems 6) Human Diseases and 7) Drug Development. Likewise, Gene Ontology is a bioinformatics tool that attempts to unify the representation of a gene across all species by classifying the gene in 3 categories which include 1) Cellular Component (CC) 2) Molecular Function (MF) or 3) Biological Process (BP). Genes that fall into the category of Cellular Component describes where in the cell a gene acts or what organelle a gene product functions in. Molecular Function deals with the elemental activities of a gene product at the molecular level, such as binding or some type of catalytic activity. Biological Process involves those genes that are important for the functioning of cells, tissues, organs, and the organism. These groupings (i.e., KEGG pathways and Gene Ontologies) would allow for a more complete understanding of molecular changes occurring during the stage of metamorphic competence of *H. elegans*.

B. Genome Sequencing of *P. luteoviolacea* and *C. lytica*

PacBio SMRT Sequencing of the *P. luteoviolacea* (HI1) genome resulted in 82,296 raw reads with a mean read length of 5,344 bp, totaling 439,749,929 nucleotides. Previous genomic sequencing of this microorganism was performed by New Mexico State University (Road Runner) using Roche 454 sequencing methods (Life Technologies) resulting in 133 contigs. To create a hybrid assembly, the Roche 454 sequencing assembly was combined with the HGAP assembly using Minimus (Sommer et al., 2007) and re-scaffolded using PacBio's AHA Program. The hybrid assembly of the Roche 454 and PacBio SMRT sequencing resulted in 10 scaffolds representing the 5.8 Mb genome of *P. luteoviolacea* (HI1).

Similarly, SMRT sequencing of the *C. lytica* (HI1) genome resulted in 156,902 raw reads with a mean read length of 5,564bp, totaling 873,038,511 nucleotides. Generated reads were then introduced into the Hierarchical Genome Assembly Process (HGAP) which includes assembly with the Celera Assembler and assembly polishing with Quiver. The complete genome of *C.*

lytica LIM-21 was used as the reference genome in the sequence assembly of *C. lytica* (HI1). The final complcte genome resulted in a single scaffold of 3,824,196 bp, in comparison to 3,765,936 bp of the published complete genome of *C. lytica* type strain LIM-21.

Both the draft genome sequcnccs of *P. luteoviolacea* and the complete genome of *C. lytica* were submitted to the RAST (**R**apid **A**nnotation of **S**ubsystem **T**echnology) server (Aziz et al., 2008) (Overbeek et al., 2013) which is a fully automated on-line servcr used to annotate complete or nearly complete bacterial and areheal genomes. RAST analysis predieted a total of 5284 coding sequences for *P. luteoviolacea* and 3340 coding regions for *C. lytica*. All coding regions for both microorganisms were then grouped into functional categories called subsystems (Figure 2, Figure 3)

IV. Expanded Accomplishments

A. Amplification of Candidate Receptor Genes and In-Situ Hybridization of Competent Larvae of *H. elegans*.

The larvae of *H. elegans* were grown to competence (day 6) and harvested on ice, fixed and stored in methanol for *in-situ* hybridization studies. Larvae that were not fixed were placed on ice and larval RNA was immediately extracted using the Qiagcn RNeasy Mini Kit, reversed transcribed, and the cDNA was generated using the SMARTer RACE cDNA Amplification kit (Clontech). Primers specific to genes found in the transcriptome database and related to bacterial recognition sequences were designed and riboprobes were synthesized. The riboprobes were then used in *in-situ* hybridization experimcnts to visualize the anatomical location of the transcripts on the larvae of *H. elegans* (Figure 1).

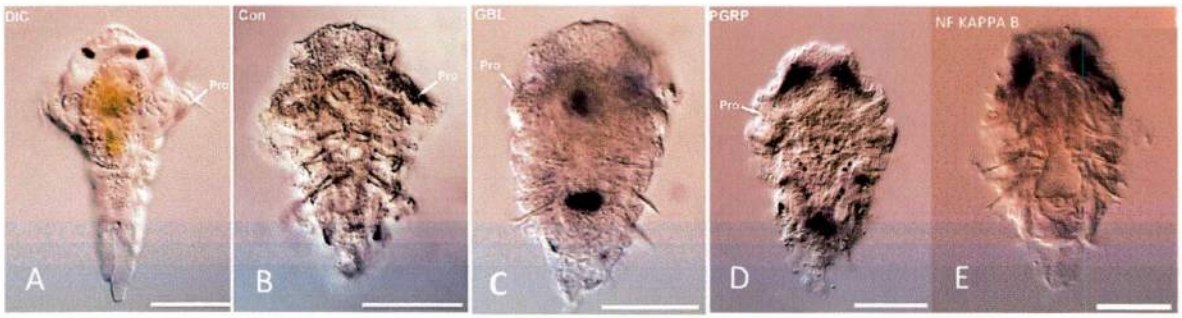

Figure 1. Whole mount in-situ hybridization of *Hydroides elegans* competent (day 5-6) larvae displaying: (A): Competent larvae, (B) Control (C) Expression of Galactose Binding Lectin (D) Expression of Peptidoglycan Recognition Protein (E) Expression of NF-KappaB

B. Differential Expression of Genes Unique to the Competent Larvae of *H. elegans*

The larvae of *H. elegans* was grown to precompetence (Day 3), competenee (day 6) and Adult stages and harvested on iee. The larval and adult RNA was then extracted using the Qiagcn RNeasy Mini Kit. The quality of RNA was then determined using the Agilent Bioanalyzer and

submitted to the Huntsman Cancer Institute at the University of Utah for transcriptome sequencing. After the alignment and assembly of the reads, all sequences were submitted to Blast and compared against databases at NCBI, Swiss-Prot, TRembl, and NR All the identified protein sequences that were differentially expressed in only competent larvae were then determined using EBSeq and called Pattern 3 (Table 1). Pattern 3 gene sequences were then analyzed using DAVID and classified into KEGG Pathways (Table 2) and Gene Ontologies (GO) (Table 3). Table 2 shows the most significant biological pathways ($p<0.05$) involved at the stage of metamorphic competence. Likewise, Table 3 shows the most significant processes involved in the GO categories of Cellular Component, Molecular Function, and Biological Process. Overall, KEGG Pathway analysis shows that at the stage of metamorphic competence there is an upregulation of genes that are involved in processing of environmental information by way of genes involved in Neuroactive Ligand Receptor and Notch Signaling Pathways. Likewise, on an organismal level, there is an up-regulation of genes involved in the development (Dorso-Ventral Axis Formation), circulatory system (Vascular Smooth Muscle), and the immune system (Complement and Coagulation Pathway). Gene Ontology results also show that genes that are significantly expressed during metamorphic competence involve receptor signaling activity (Biological Process – GPCR, Molecular Function – Neurotransmitter Activity).

Table 1. Various pattern designations for the differential expression of genes of precompetent, competent and adult stages of *H. elegans*.

Pattern	Gene Expression	Precompetent	Competent	Adult
1	Same for all stages	1	1	1
2	Unique to precompetent stage, common to competent and adult stages	2	1	1
3	Unique to competent stage, common to precompetent and adult stages	1	2	1
4	Unique to adult stage, common to precompetent and competent stages	2	2	1
5	Different in all stages	3	2	1

Table 2. KEGG Pathways involved at the stage of competence for *H. elegans*

KEGG Pathway	Category	P-value
Neuroactive Ligand Receptor Interaction	Environmental Information Processing	9×10^{-5}
Vascular Smooth Muscle	Organismal	6×10^{-3}
Notch Signaling Pathway	Environmental Information Processing	2.4×10^{-2}
Complement and Coagulation Cascade	Organismal	2.4×10^{-2}
Proteosome	Genetic Information Processing	2.5×10^{-2}
Dorso-Ventral Axis Formation	Organismal	3.8×10^{-2}
Chondroitin Sulfate Biosynthesis	Metabolism	4.6×10^{-2}

Table 3. Overview of Gene Ontology Classification for Competent Larvae

Gene Ontology	Process	P-value
Cellular Process	Integral to Plasma Membrane	2.4×10^{-15}
	Extracellular Region	8.8×10^{-8}
	Synapse Region	5.4×10^{-7}
Molecular Function	Calcium Ion Binding Activity	4.4×10^{-17}
	Guanylate Cyclase Activity	1.5×10^{-12}
	Neurotransmitter Receptor Activity	6.9×10^{-10}
Biological Process	G-Protein Coupled Receptor Signaling Pathway	6.8×10^{-8}
	cGMP Biosynthesis	5.7×10^{-13}
	Carotenoid metabolism	1.7×10^{-2}

C. RAST Analysis of *P. luteoviolacea* and *C. lytica* Genomes

A recent study on *P. luteoviolacea* has found that the inductive gene products responsible for the settlement and metamorphosis of *H. elegans* are components of a phage tail-like bacteriocin (Shikuma et al., 2014). RAST analysis of the *P. luteoviolacea* genome (Figure 2) shows that there are 14 genomic elements within the genome of *P. luteoviolacea* that are found in the category Phages, Prophages, Transposable Elements and Plasmids. Most of these sequences encode for phage structural components such as the sheath, tube and base plate of a phage tail. This is in contrast to the *C. lytica*, the second inductive strain, which has 2 elements related to Phages, Prophages, Transposable element and Plasmids contained within its genome as determined by RAST (Figure 3). The two genomic elements of *C. lytica* encode for a phage-tail fiber protein and a Listeria Pathogenicity Island, but no other phage-tail structural component is identified (i.e., phage sheath, tube or base plate). Thus, it is appears that the cue responsible the settlement and metamorphosis of *H. elegans* produced by *C. lytica* involves a different mechanism and requires further analysis and research in the future.

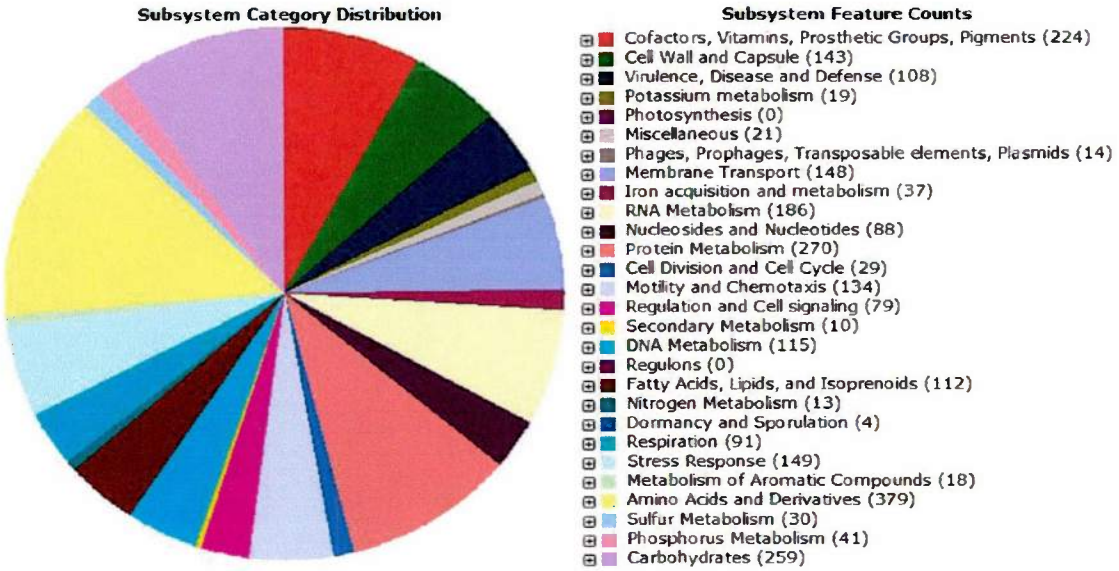

Figure 2: RAST analysis of *Pseudoalteromonas luteoviolacea*

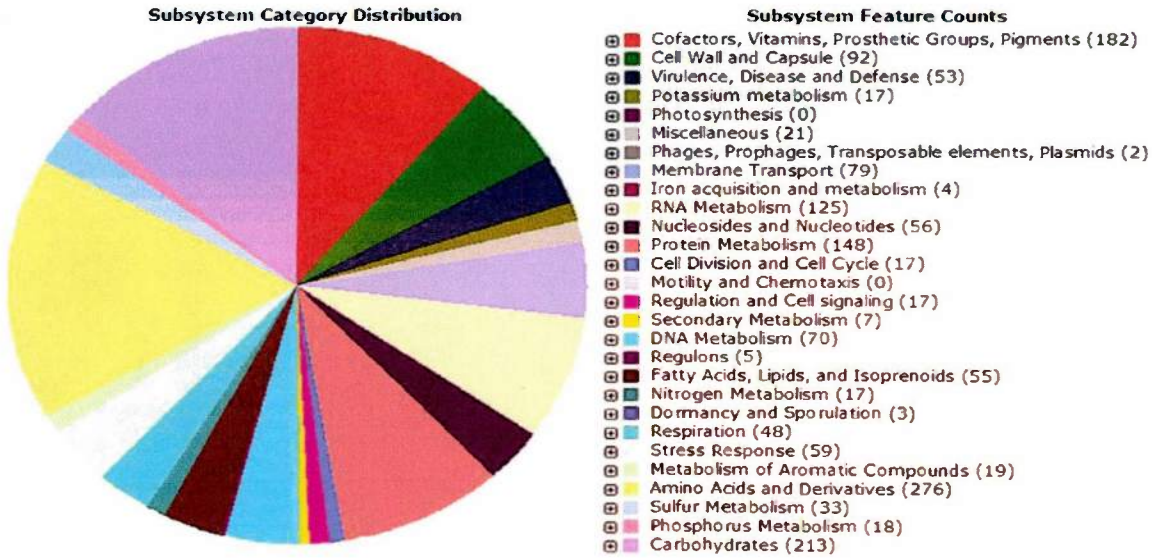

Figure 3. RAST Analysis of *Cellulophaga lytica*

V. Work Plan

This is the Final Report

VI. Major Problems/Issues

None are anticipated at this time.

VII. Technology transfer.

We participate in no activities that involve technology transfer.

VIII. Foreign collaborations and supported foreign nationals.

We supported no foreign collaborations.

IX. Productivity.

Papers Published/to be published on this Grant.

Huang, Y., S. Callahan and M. G. Hadfield. 2012. Recruitment in the sea: bacterial genes required for inducing larval settlement in a marine worm. Scientific Reports 2:228 | DOI: 10.1038/srep00228.

Shikuma, N.J., Pilhofer, M., Weiss, G.L., Hadfield, M.G., Jensen, G.J., Newman, D.K. 2014. Marine tubeworm metamorphosis induced by arrays of bacterial phage tail-like structures. *Science* 343(6170):529-533.

Hadfield, M.G., B. Nedved, S. Wilbur and M. A. R. Koehl. 2014. Biofilm cue for larval settlement in *Hydroides elegans* (Polychaeta): is contact necessary? Marine Biology (in review).

Hadfield, M. G., A. Asahina, S. Hennings and B. Nedved. 2014. The bacterial basis of biofouling. Indian Journal of Geo-Marine Sciences. (in press).

Asahina, A.Y. and Hadfield, M.G. 2014. *De Novo* Transcriptome Sequencing of the Serpulid Polychaete *Hydroides elegans*. *PlosOne*. In Prep.

Asahina, A.Y. and Hadfield, M.G. 2014. Draft Genome Sequence of *Pseudoalteromonas luteoviolacea* Hawaii Strain. *Genome Announcement*. In prep.

Asahina, A.Y. and Hadfield, M.G. 2014. The Complete Genome Sequence of *Cellulophaga lytica* Hawaii Strain. *Genome Announcement*. In prep.

Asahina, A.Y., and Hadfield, M.G. 2014. Comparative analysis of two *Cellulophaga lytica* genomes. In Prep.